

Disparities in dermatology AI performance on a diverse, curated clinical image set

Year: 2021 | Citations: 323 | Authors: R. Daneshjou, Kailas Vodrahalli, Weixin Liang, R. Novoa, Melissa Jenkins

Abstract

An estimated 3 billion people lack access to dermatological care globally. Artificial intelligence (AI) may aid in triaging skin diseases and identifying malignancies. However, most AI models have not been assessed on images of diverse skin tones or uncommon diseases. Thus, we created the Diverse Dermatology Images (DDI) dataset—the first publicly available, expertly curated, and pathologically confirmed image dataset with diverse skin tones. We show that state-of-the-art dermatology AI models exhibit substantial limitations on the DDI dataset, particularly on dark skin tones and uncommon diseases. We find that dermatologists, who often label AI datasets, also perform worse on images of dark skin tones and uncommon diseases. Fine-tuning AI models on the DDI images closes the performance gap between light and dark skin tones. These findings identify important weaknesses and biases in dermatology AI that should be addressed for reliable application to diverse patients and diseases.